Level dependency of auditory stream segregation

Hauth, Christopher; Christiansen, Simon Krogholt; Dau, Torsten

Published in:
Proceedings of the Deutsche Gesellschaft für Akustik - 2013

Publication date:
2013

Level Dependency of Auditory Stream Segregation

Christopher Hauth[1], Simon Krogholt Christiansen[2], Torsten Dau[2]

[1] *Jade Hochschule Oldenburg, 26121 Oldenburg, Germany, Email: christopher.hauth@ewetel.net*

[2] *Centre for Applied Hearing Research, DTU, 2800 Lyngby, Denmark*

Introduction

In everyday life, sound entering the ear is often composed of a mixture of sounds emitted by different acoustic sources in the environment. Despite this, the human auditory system is capable of segregating this mixture of sounds into components or streams according to the acoustical source by which each sound is emitted. This enables a listener to direct his/her attention towards a single acoustical source, which is referred to as auditory stream segregation [1].

The ability to segregate sound sources is thought to be a combination of bottom-up processing driven by acoustic cues, as well as top-down schemata-based processing (e.g. [1]). One of the acoustic cues important for the bottom-up processing of auditory stream segregation is frequency separation. Sounds that have similar frequency content are more likely to be grouped together into a single auditory stream than sounds that are spectrally well separated. This phenomenon was investigated in detail by [2], by presenting simple stimuli consisting of two tones, A and B, in an ABA-ABA pattern, as shown in Figure 1. Depending on the tone-repetition time (TRT), the time between onsets of successive tones, and the frequency separation of the tones (Δf), the stimulus evoked one of two different percepts: Either the tones fused together into a single stream and a characteristic "galloping rhythm" was heard, or the tones split into two streams and the listener's attention was drawn either to the slowly repeating B---B---B tone stream, or the fast repeating A-A-A-A stream.

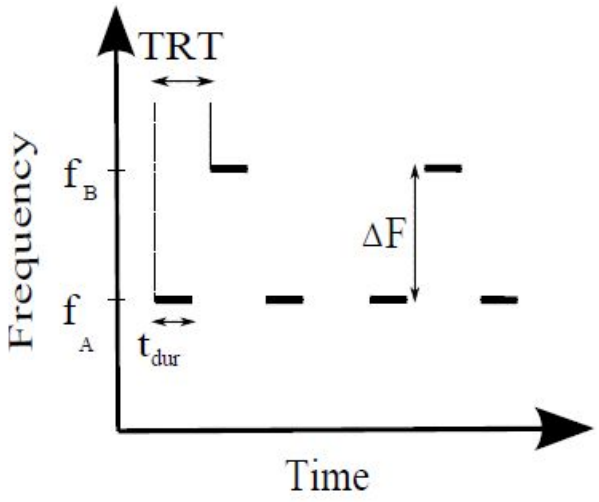


Figure 1: Stimulus used by [2] to investigate sequential grouping. The stimulus consisted of two pure tones, A and B, presented in an ABA-ABA pattern. The time between the onsets is given by TRT and the frequency separation between the tones by Δf in semitones.

By instructing the listeners to either hold on to the galloping rhythm, or to try to hold on to one of the two tones [2], two boundaries were obtained, as shown in Figure 2: The temporal coherence boundary (TCB) indicates the maximum frequency separation where it is possible to perceive two tones as a single coherent stream, and the fission boundary (FB) is the smallest frequency separation where it is possible to perceive the stimulus as two separate streams and selectively attend to a single of them.

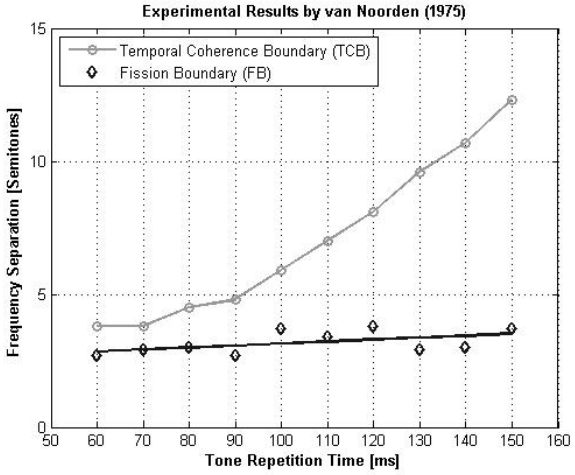


Figure 2: Temporal Coherence Boundary (TCB) and Fission Boundary (FB) determined by [1]. The TCB increases as a function of tone repetition time (TRT), whereas the FB is constant over TRT.

Several theories have been proposed to account for the findings of [2]. One of the concepts assumes that sounds that excite well separated places on the basilar membrane (i.e., different tonotopic positions) are perceived as two streams, whereas sequences with overlapping excitation are grouped into one stream, i.e. perceived as emitted by the same source. This concept is often referred to as the "peripheral channelling theory" (e.g. [3]). This theory is to some extent supported by results from physiological measurements of auditory stream segregation in song-birds [4], which furthermore suggest that physiological forward masking may be responsible for the effect of TRT on the TCB. For fast-repeating tones (small TRT), forward masking may reduce the spread of excitation along the tonotopic axis. This will, in turn, cause a reduced tonotopic overlap, leading to a segregated percept. Conversely, for slowly repeating tones (large TRTs) the interval between successive tones is long enough for the effects of forward masking to diminish, and the percept then mainly depends on the bandwidth of the auditory filters.

In the present study, the "peripheral channelling" hypothesis was tested by measuring the TCB at overall levels of 40, 60 and 80 dB SPL. Due to the level-dependent frequency selectivity of the auditory system, higher levels give rise to

broader excitation patterns. If overlapping excitation plays a dominant role, stream segregation should thus occur at larger frequency separations for higher levels than for lower levels.

Perceptual experiments based on the stimuli of [2] were designed and conducted by three normal-hearing listeners.

The same stimuli were used as the input to a model of auditory stream segregation. The model is based on the computational auditory signal processing and perception (CASP) model [5], simulating the signal processing of the peripheral human auditory system, combined with a temporal coherence analysis proposed by [6]. This combined model was suggested by [7] and has been shown to quantitatively account for the data of [2] shown in Figure 1.

Model description

The preprocessing stage realised by CASP consists of an outer- and middle-ear filter, a basilar membrane filterstage, a haircell transduction stage and an adaptation stage.

The outer- and middle ear filter are designed as linear phase FIR-filter (order=512). The filtering on the basilar membrane is realised by a dual resonance nonlinear (DRNL) filterbank [8]. In the DRNL-stage, the signal is processed independently in two paths, a linear and a nonlinear path. In the linear path, a linear gain is applied before the signal is bandpass-filtered by a cascade of two second-order gammatone filters, followed by a cascade of four second-order lowpass filters. In the nonlinear path, the signal is filtered by a cascade of two second-order gammatone filters, followed by a broken-stick nonlinearity. Finally, the signal is again filtered by a cascade of two second-order gammatone filters and a second-order lowpass filter. The output of the DRNL filterbank is the sum of the linear and the nonlinear path. The filterbank ranged from 300 to 3000 Hz with one equivalent rectangular bandwidth (ERB) [9] spacing. The DRNL filterbank is followed by a stage simulating hair-cell transduction, realised by a half wave rectification and a second-order butterworth lowpass filter with a cut-off frequency of 1 kHz. An expansion device, realised as an instantaneous squaring operation and mirroring the square-law behaviour of neural firing rate [10] is included, before the signal is fed to the adaptation stage. Neural adaptation is simulated by a series of five feedback loops, with time-constants of 5, 50, 129, 253 and 500 ms [11]. As a last step, the signal is filtered with a first-order lowpass filter with a cut-off frequency of 8 Hz, simulating the limited temporal resolution of the auditory system. This stage corresponds to the temporal integration stage[6].

A temporal coherence analysis, as proposed by [6] was used as back-end for the model of auditory stream segregation. A coherence matrix for the peripheral frequency channels of the auditory spectrogram after temporal integration is computed. This is done for every discrete time step, before the coherence matrices are averaged over time. Finally, an eigenvalue decomposition is performed on the coherence matrix in order to determine the number of significant eigenvalues. The stimuli used in this experiment produced either a one-stream or a two-stream-percept, and by calculating the ratio between the second-largest (λ_2) and the largest eigenvalue (λ_1) the strength of a two-stream-percept could be estimated. A low eigenvalue ratio indicates a one-stream-percept and with a two-stream-percept becomes more likely with increasing ratio.

Methods

The experiment used a constant stimuli paradigm, where the stimulus consisted of an ABA-tone pattern as used in [2]. The ABA pattern is shown schematically in Figure 1. The A-tone was fixed at a frequency of 1 kHz and the B-tone was varied between –13 to 13 semitones relative to the A-tone in the following steps: [-13, -11, -9, -7, -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5, 7, 9, 11, 13]. The starting frequency and the direction of change of the B-tone were randomized to minimize effects of direction of change. The ABA-pattern consisted of pure tones with durations of 60 ms, including 5 ms raised-cosine ramps.

The TCB was determined for six TRTs of 70, 90, 110, 130, 150 and 170 ms at three different overall level of 40, 60 and 80 dB SPL. The listeners were instructed to listen for a coherent stream by focusing on the galloping rhythm as soon as possible and hold on to this percept as long as possible. To make sure that all subjects became familiar with the galloping rhythm, a short presentation of the sequence at a frequency separation of 1 semitone and a TRT of 70 and 130 ms was provided, which reliably produced a one-stream-percept. For a presentation of a two-stream-percept, the frequency separation was increased to 13 semitones.

All stimuli were generated in MATLAB (MathWorks) using the AFC-Toolbox, developed at the Technical University of Denmark (DTU) and Carl von Ossietzky University, Germany, and presented monaurally to the subject's left ear via a RME DIGI 96/8 PAD soundcard and HD 580 Sennheiser headphones.

A Yes/No-experiment was used to determine the TCB. Each level and TRT was tested ten times, resulting in 180 runs. Overall, six sessions were run, each lasting 1h including breaks. The experiments were conducted in a sound attenuated and electrically shielded booth. The listeners responded via the keyboard of a PC whereby no visual feedback was provided. Three normal-hearing listeners (23, 23 and 24 years old) participated in the experiment, including the first author of this manuscript. All listeners had previous experience in psychoacoustic experiments and subject FG was paid for his effort.

Results

The average results across listeners are presented in Figure 3. The data were roughly symmetrical around $\Delta f = 0$ semitones. Thus, for easier comparison for each TRT, the frequency separation $|\Delta f|$ was averaged, similar to the processing performed by [2]. Consistent with the original data from [2], the TCB increases as a function of TRT. However, in contrast to our hypothesis, the data shows no effect of level. The averaged results for all subjects were used to investigate if there are any significant differences between the TCB across level. The TRT of 170 ms was excluded from the analysis, because of ceiling effects. First,

a Lilliefors' test was conducted. A level of significance of α=0.01 was chosen. It revealed that the data were not normally distributed. A Wilcoxon signed rank test with a level of significance of α=0.01 was chosen for statistical analysis. The statistical test showed no significant differences in the TCB across level for the average results.

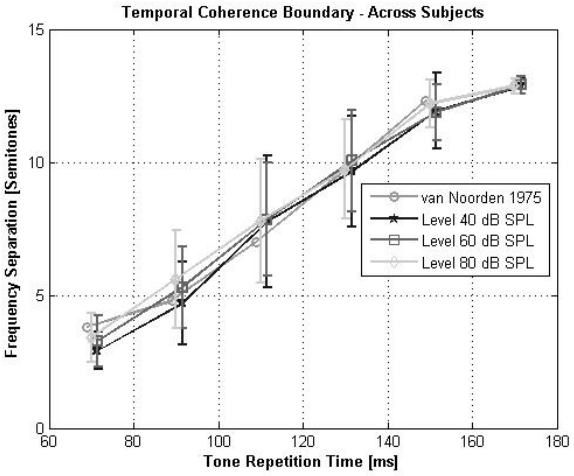
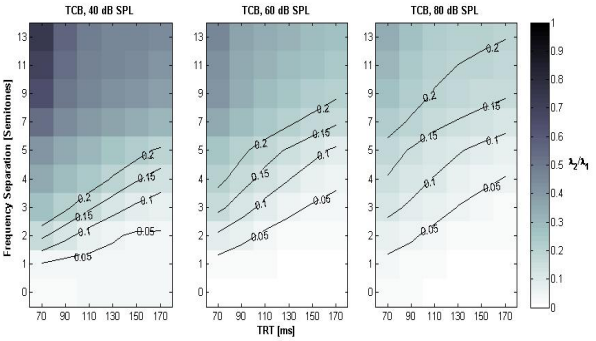


Figure 3: Measured mean and standard error of the TCB. The black stars show the result for 40 dB SPL, the dark gray squares for 60 dB SPL and the bright gray diamonds for 80 dB SPL. For comparison, the results of [2] are indicated by the gray circles.

The corresponding simulations are shown in Figure 4. The eigenvalue-ratios for each frequency separation and TRT are shown, indicated by a colorbar ranging from 0 (white) to 1 (black), corresponding to a most likely one-stream-percept (white) and a most likely two-stream-percept (black). Some contours for eigenvalue ratios of 0.05, 0.1, 0.15 and 0.2 are shown explicitly. Consistent with the data, the model predicts a one-stream percept for low frequency separations, and an increasing likelihood of a two-stream percept for higher frequency separations. Similar to the measured TCB, the model also shows that a larger frequency separation is needed to produce a two-stream percept for large TRTs than for small TRTs. However, in contrast to the data, the simulation shows a clear effect of stimulation level. With increasing sound pressure level, the eigenvalue-ratio-contours shifts towards larger frequency separations.



a) Level: 40 dB SPL b) Level: 60 dB SPL c) Level: 80 dB SPL

Figure 4: Out of the model of auditory stream segregation, using the same stimuli as for the measurements mentioned above. The colour indicates the eigenvalue ratio, and the black lines show the eigenvalue contours at 0.05, 0.1, 0.15 and 0.2.

Discussion

The experimental results showed no significant effect of level, despite the 40 dB of dynamic range of the stimuli. This is in contrast to the concept that overlapping excitation plays a dominant role in primitive stream segregation, and may indicate that more central stages are involved in determining the perceptual organization of spectral components, making the percept robust against changes in level.

The simulation showed a substantial effect of level, consistent with the peripheral channelling hypothesis. By increasing the overall level, the excitation pattern evoked by each tone becomes broader. Therefore, overlapping excitation occurs even at larger frequency separations and the predicted TCB is shifted towards larger frequency separations.

The differences between perceptual data and simulations might be explained by the fact that the model only takes bottom-up processes of the peripheral hearing into consideration and no central stages (such as, e.g., a level-dependent criterion for a two stream percept), which might be responsible for the "robust" results across level.

References

[1] Bregman, A. S.: Auditory Scene Analysis – The Perceptual Organization of Sound.

A Bradford Book, Cambridge, Massachusetts, 1990

[2] van Noorden, L. (1975) : Temporal Coherence in the Perception of a Tone Sequence, Ph.D. thesis, The Netherlands, Tech. Hogeschool, Eindhoven, Institute of Perception Research

[3] Hartmann, W. M., Johnson, D.: Stream Segregation and Peripheral Channeling.
Music Perception: An interdisciplinary Journal **9(2)** (1991), 155-183

[4] Bee, M. A. and Klump, G. M.: Primitive Auditory Stream Segregation: A Neurophysiological Study in the Songbird Forebrain.
Journal of Neurophysiology **92** (2004), 1088-1104

[5] Jepsen, M. L., Ewert, S. D., and Dau, T.: A computational model of human auditory signal processing and Perception.
The Journal of the Acoustical Society of America **124(1)** (2008), 422-438

[6] Elhilali, M., Ma, L., Micheyl, C., Oxenham, A. J., and Shamma, S. A.: Temporal Coherence in the Perceptual Organization and Cortical Representation of Auditory Scenes.
Neuron **61(2)** (2009), 317-329

[7] Christiansen, S. K., Jepsen, M. L., and Dau,T.: A physiologically inspired model of auditory stream segregation based on temporal coherence analysis.

Proceedings of Meetings on Acoustics **15** (2012), p. 050001

[8] Lopez-Poveda, E. A. and Meddis, R.: A human nonlinear cochlear filterbank.
The Journal of the Acoustical Society of America **110(6)** (2001), 3107-3118

[9] Glasberg, B. R., Moore, B. C.: Derivation of auditory filter shapes from notched-noise data. Hearing Research **47** (1990), 103-138

[10] Yates, G. K., Winter, I. M., and Robertson, D.: Basilar membrane nonlinearity determines auditory nerve rate-intensity functions and cochlear dynamic range. Hearing Research **45(3)** (1990), 203-219

[11] Dau, T., Püschel, D., and Kohlrausch, A.: A quantitive model of the "effective" signal processing in the auditory system. I. Model structure.
The Journal of the Acoustical Society of America **99(6)** (1996), 3615-3622